★★★★★

Predicting Hotel Booking Cancellation

Sarah Alfrihidi

Contents

1

Problem Definition

Project Problem

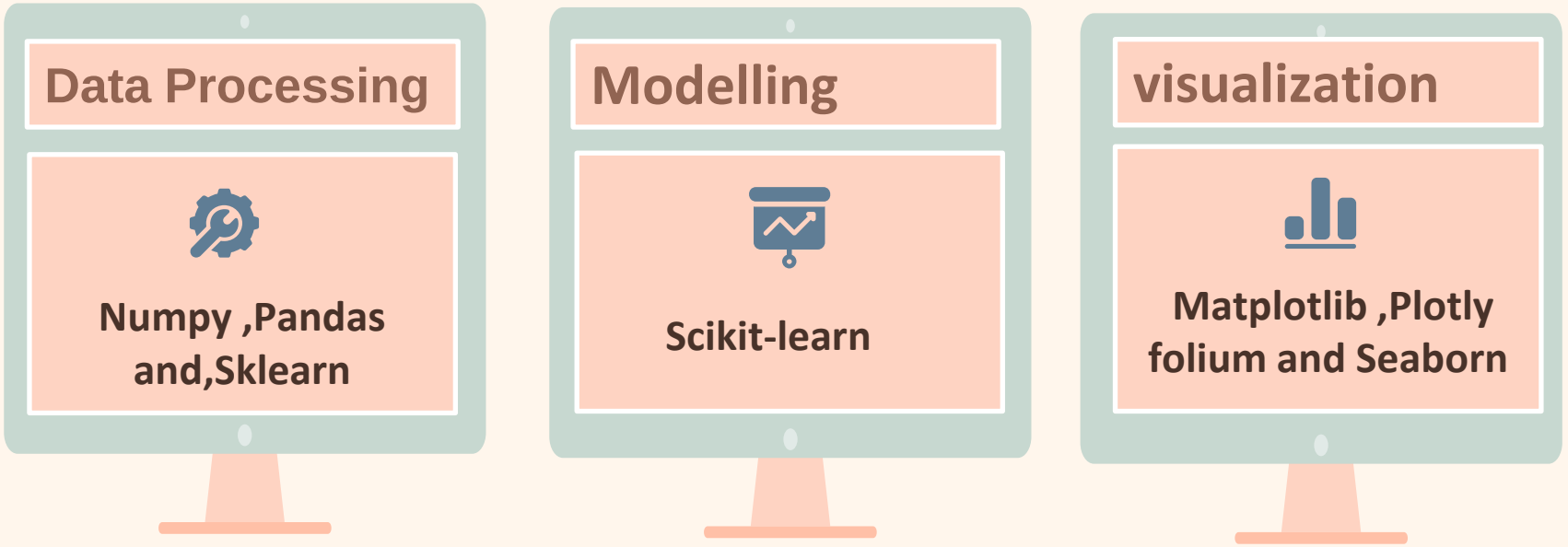
To predict whether a hotel booking will be cancelled using current and prior booking information.

Why is this important?

Throughout the year, hotels get numerous bookings, and many of these bookings are canceled. Cancelled bookings result in a loss of revenue for the company; if management could foresee whether a hotel reservation will be canceled, it would be in their best interests to assign the room to someone else and win their business. This research seeks to use current and historical booking data to predict if a hotel reservation will be canceled.

2

Tools


Data Processing
Numpy ,Pandas and,Sklearn
Modelling
Scikit-learn
visualization
Matplotlib ,Plotly folium and Seaborn

3

Dataset

Hotel booking Dataset

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	meal	country
0	Resort Hotel	0	342	2015	July	27	1	BB	PRT
1	Resort Hotel	0	737	2015	July	27	1	BB	PRT
2	Resort Hotel	0	7	2015	July	27	1	BB	GBR
3	Resort Hotel	0	13	2015	July	27	1	BB	GBR
4	Resort Hotel	0	14	2015	July	27	1	BB	GBR
...	...	...	...	...	...	...	...	...	...
119385	City Hotel	0	23	2017	August	35	30	BB	BEL
119386	City Hotel	0	102	2017	August	35	31	BB	FRA
119387	City Hotel	0	34	2017	August	35	31	BB	DEU
119388	City Hotel	0	109	2017	August	35	31	BB	GBR
119389	City Hotel	0	205	2017	August	35	29	HB	DEU

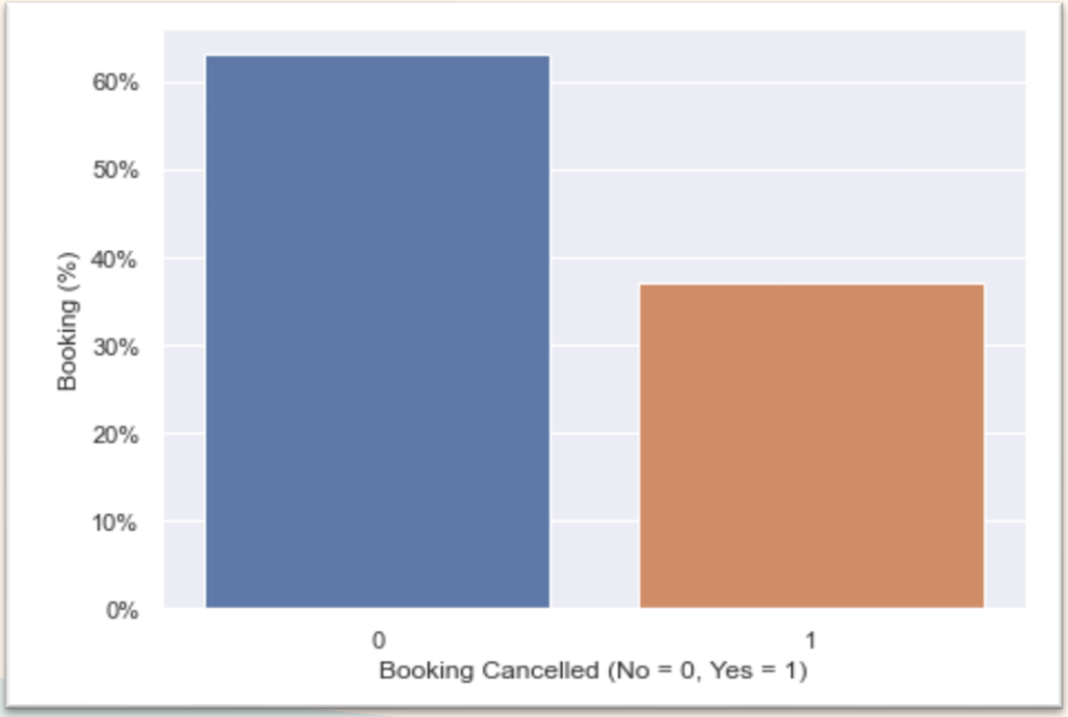
This data set includes a single file that compares various booking details between two hotels : city hotel and resort hotel.

Findings

4

Plot the count plot for both the hotel types

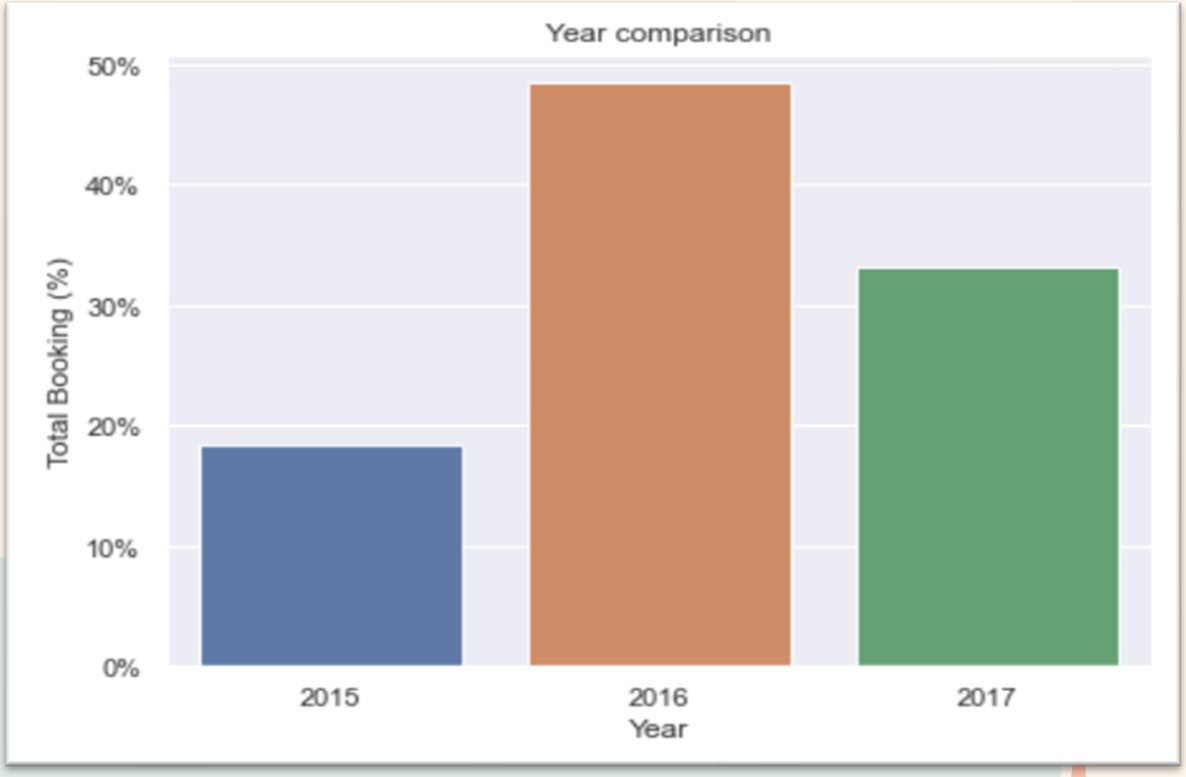
More than 60% of the population booked the City hotel



Bookings got canceled 37% of the time. While booking guests did checks-in (did not cancel the booking) almost 63% of the time.

How Many Bookings Were Cancelled?

What Is The Percentage Of Booking For Each Year?

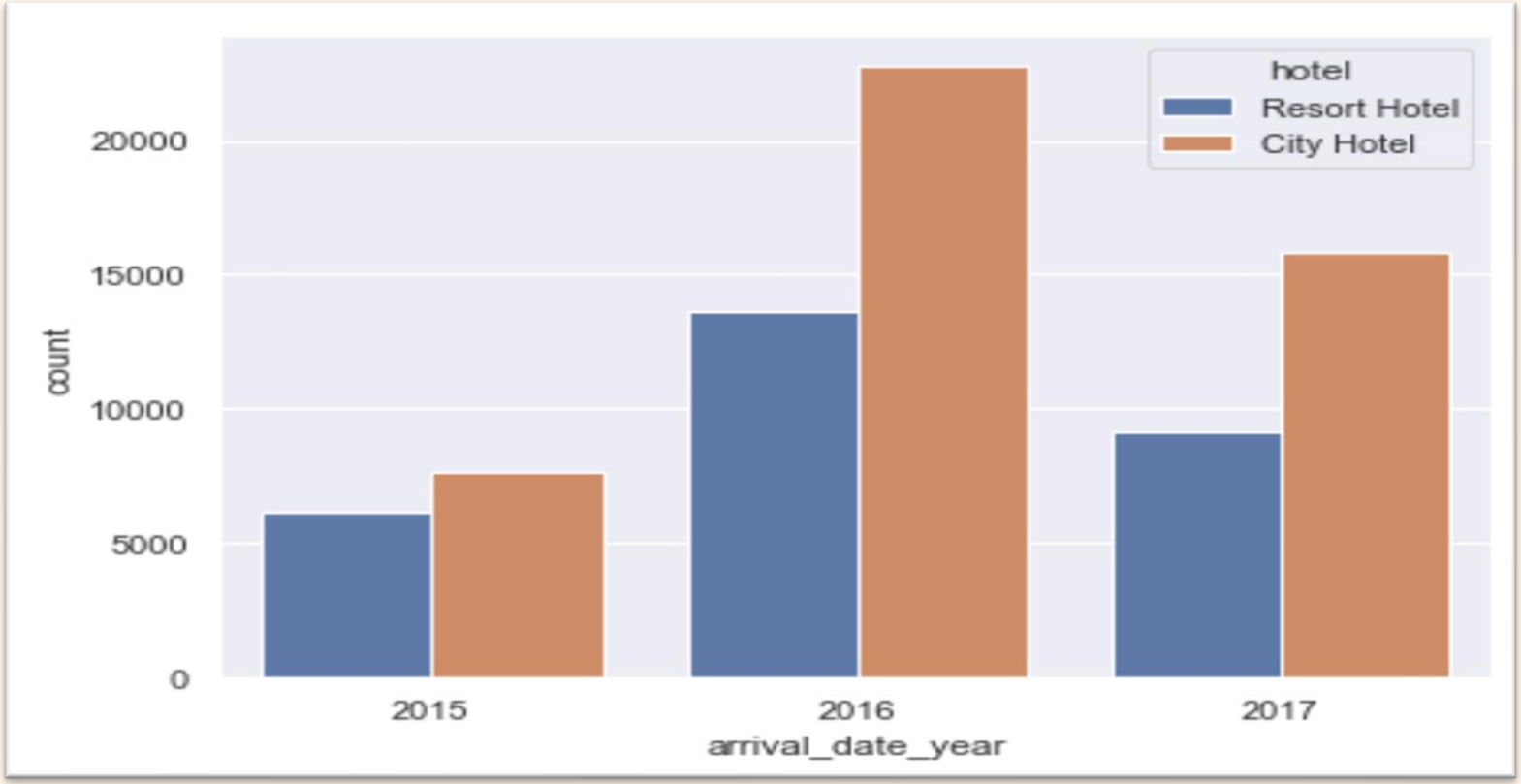


More than double bookings were made in 2016,compared to the previous year.

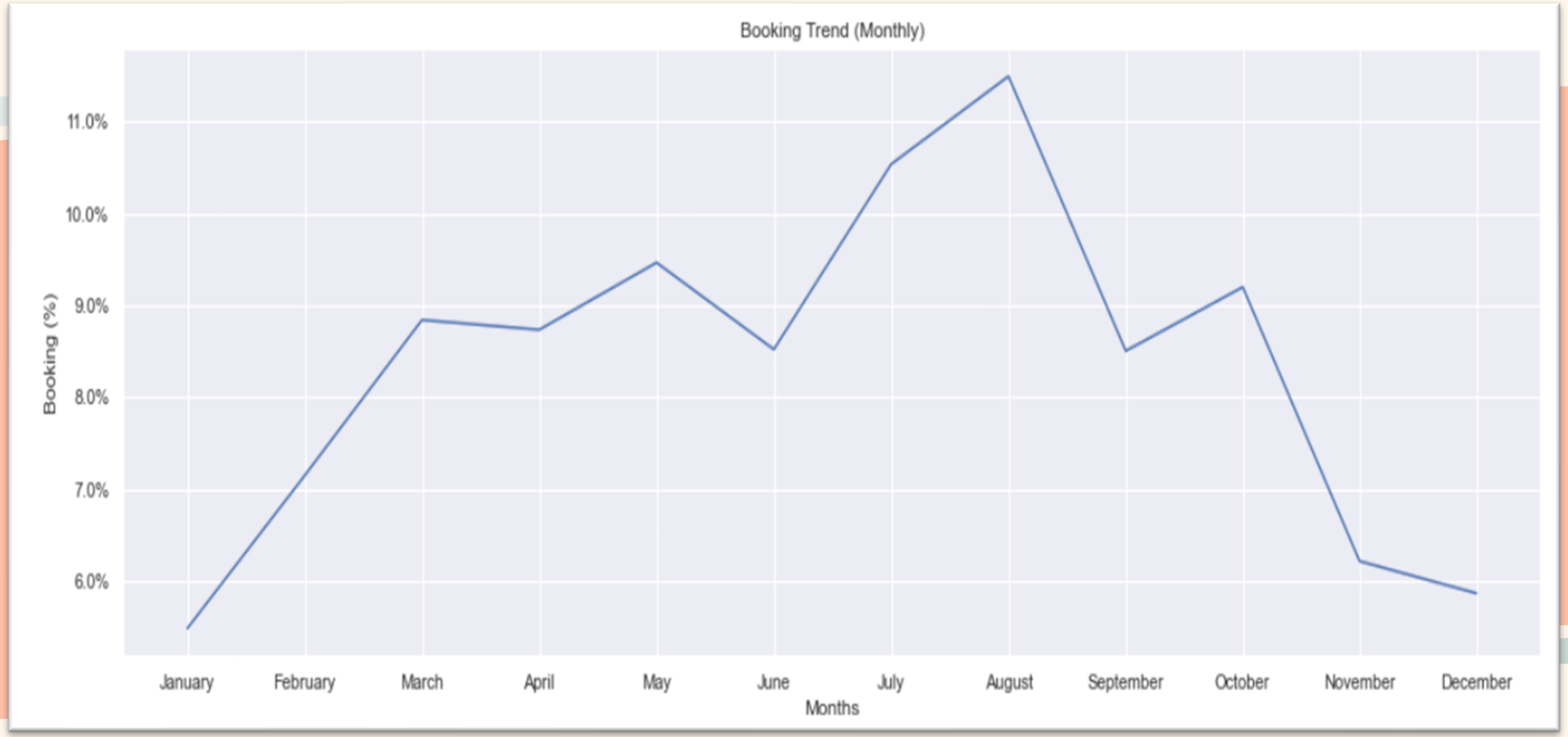
But the bookings decreased by almost 15% the next year.

★★★★★

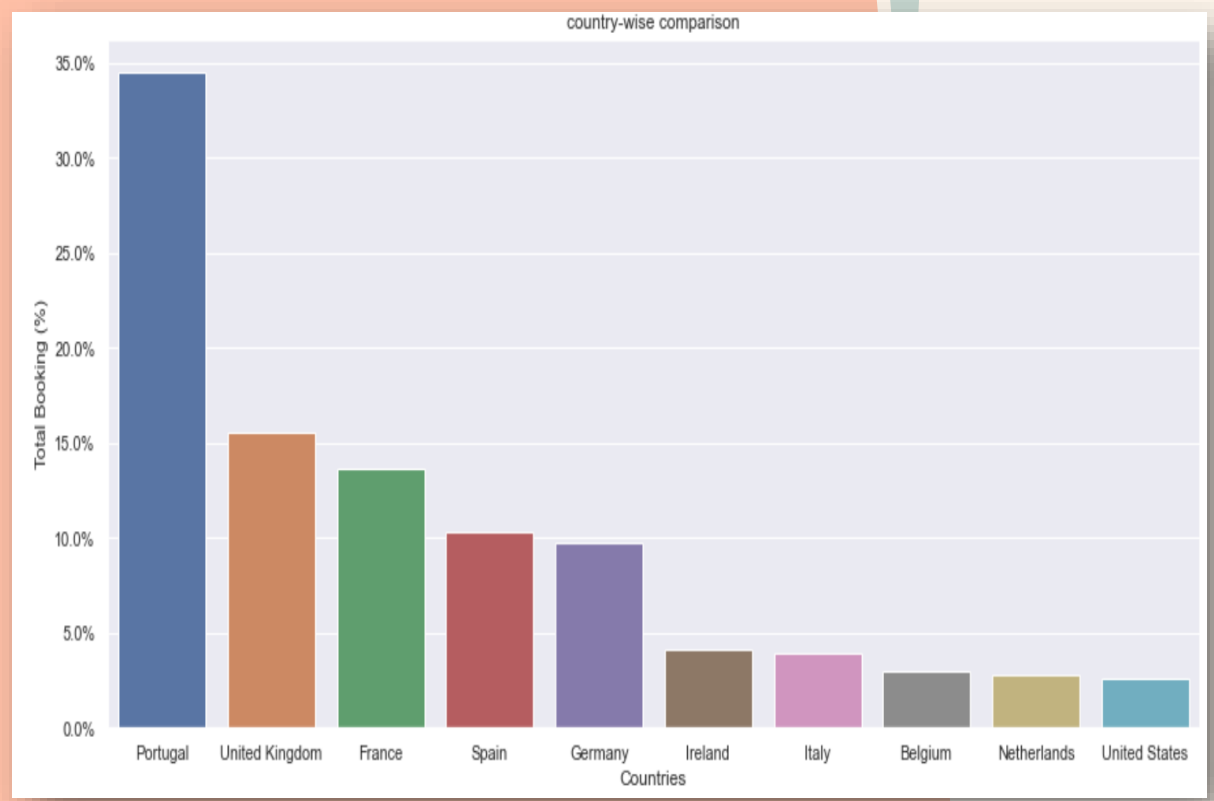
Separate By The Hotel Type



Which Is The Busiest Month For Hotels?

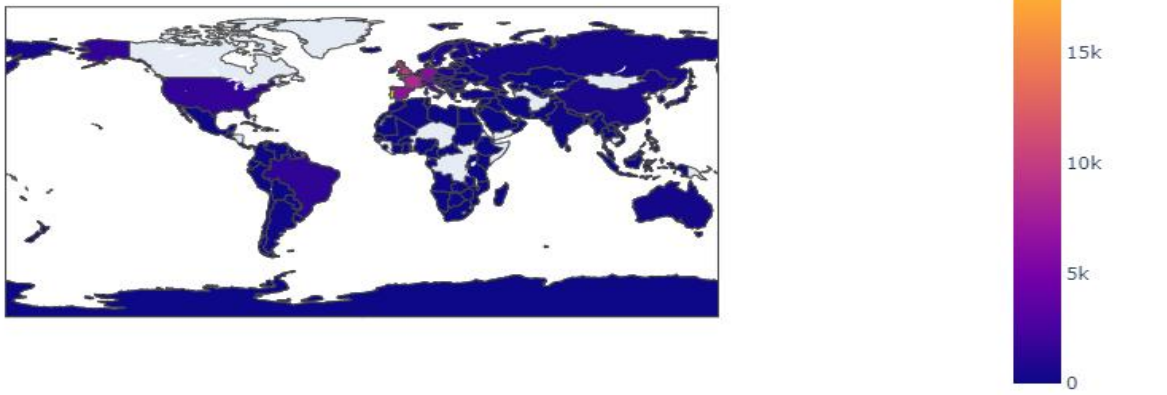


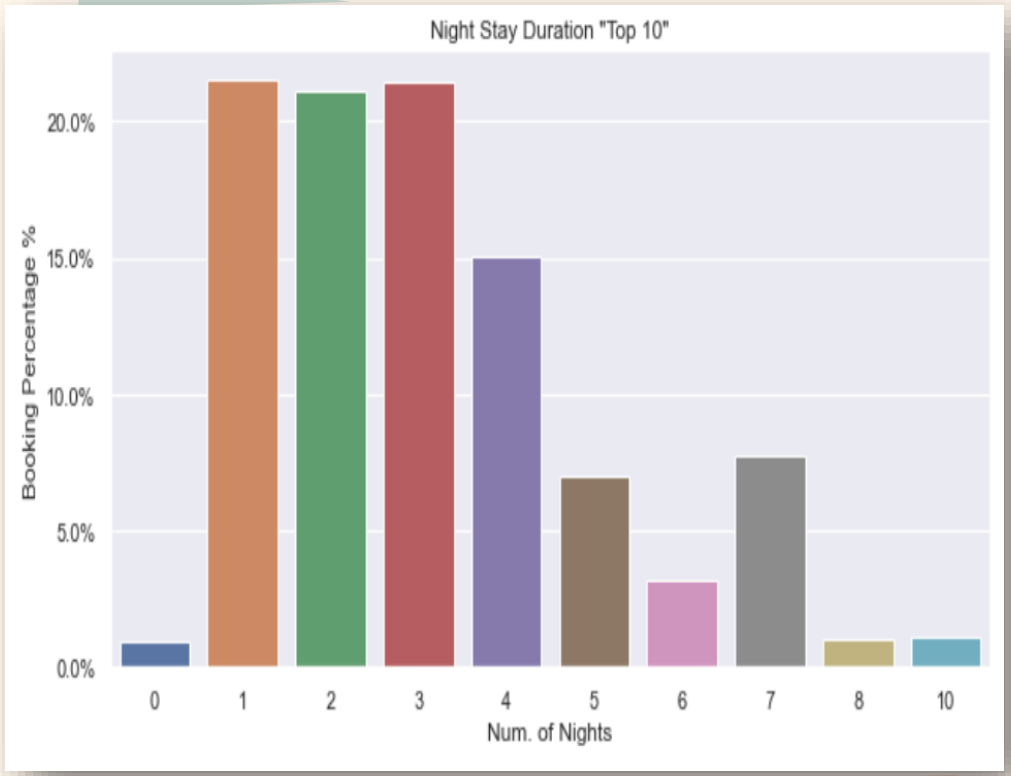
The most guests are coming ??



Portugal, UK and France, Spain and Germany are the top countries from most guests come, more than 80% come from these 5 countries.

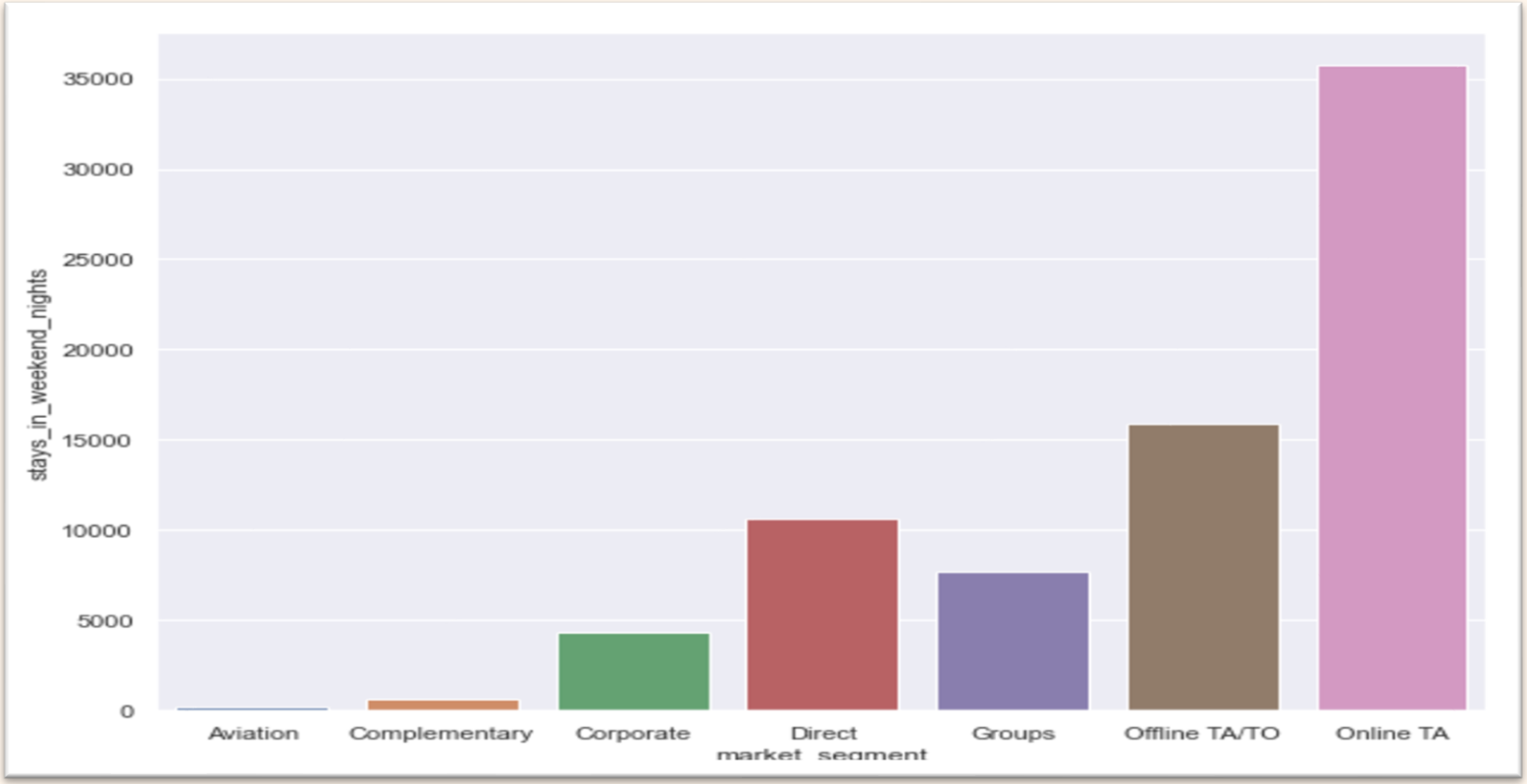
The most guests are coming display in map



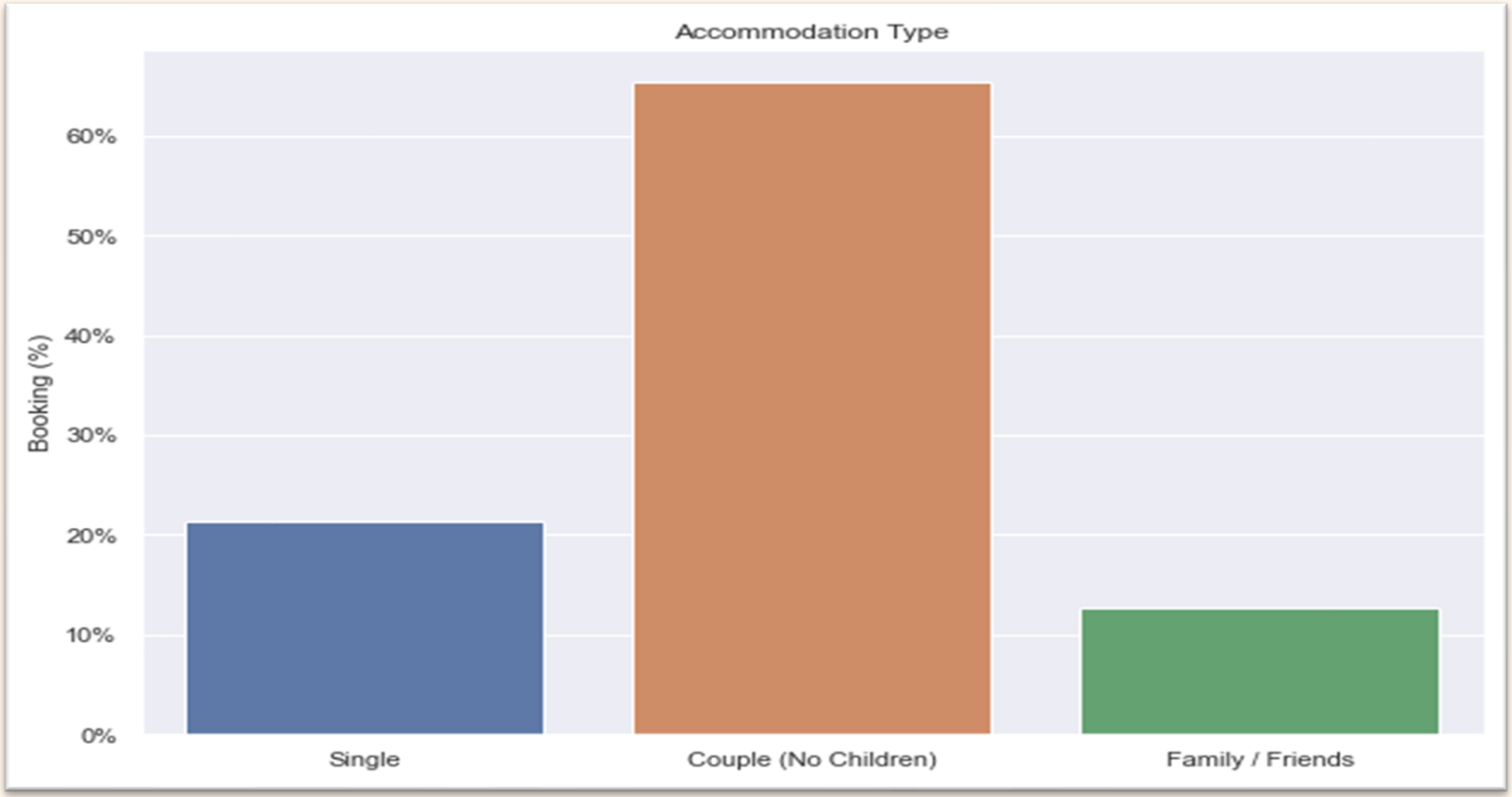


How Many Days Do People Stay In A Hotel?

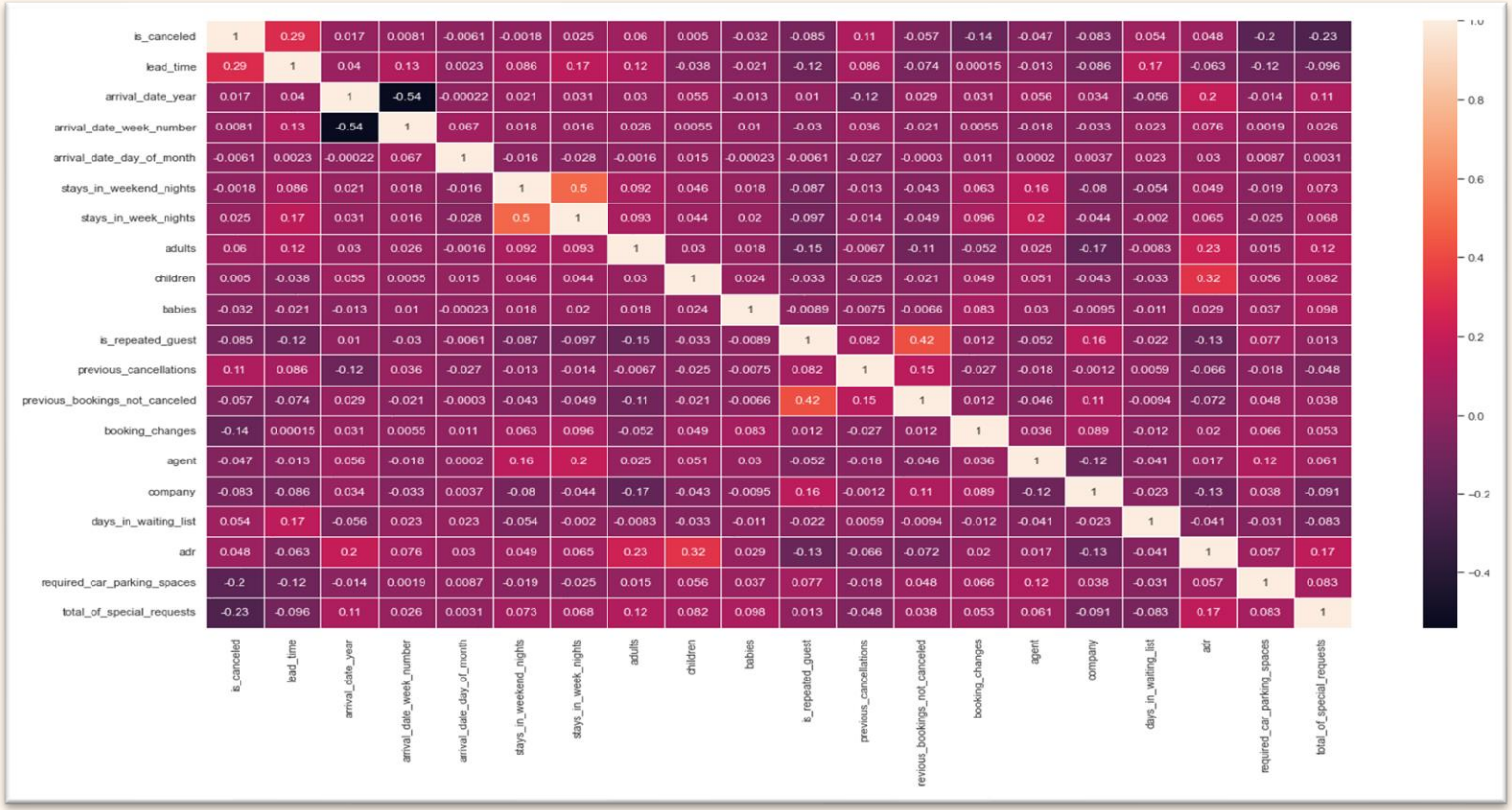
Online agents are the best market for reservation by people



Customer Type Single, Couple, Multiple Adults, And Family



Display Heatmap



05

Data Model

❑ Logistic Regression Model

	Not Canceled = 0	Canceled =1
precision	0.98	0.96
Recall	0.99	0.94
F1-score	0.98	0.95

❑ KNN Model

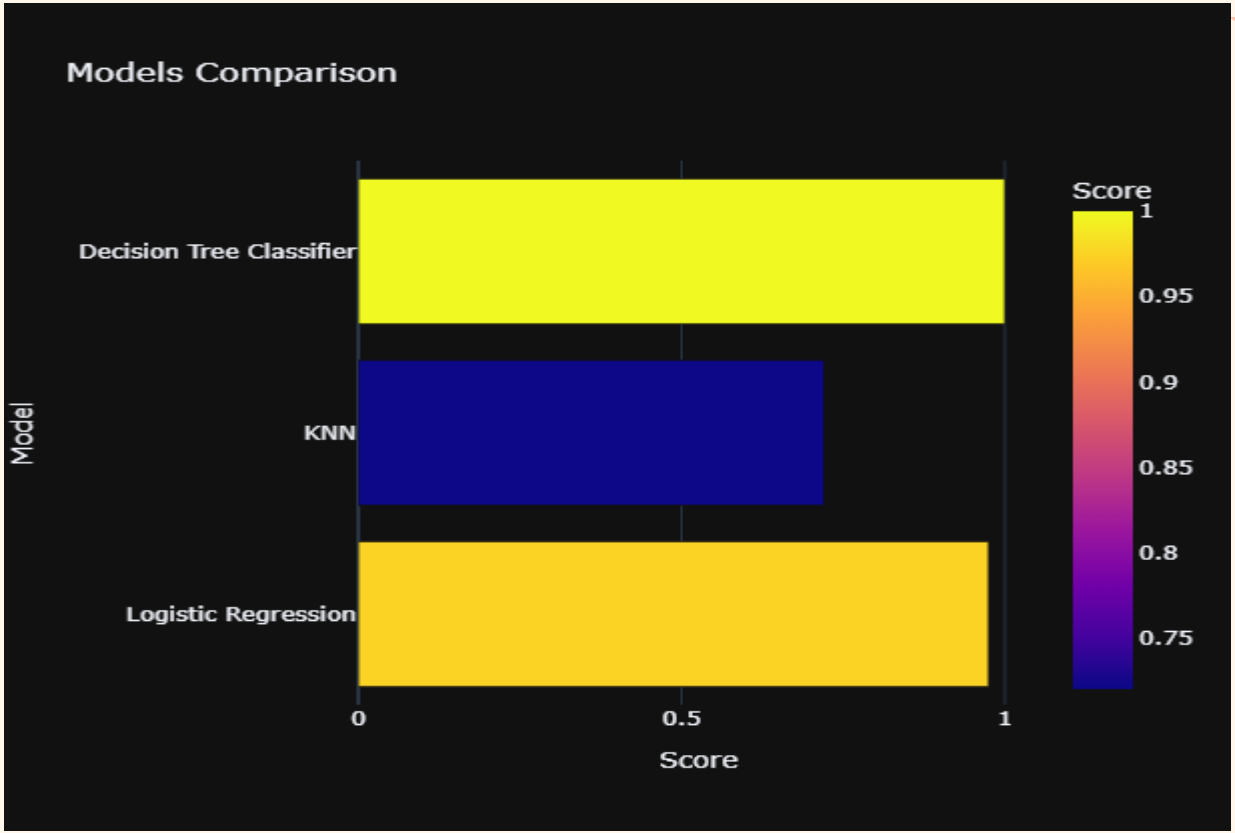
	Not Canceled = 0	Canceled =1
precision	0.78	0.48
Recall	0.87	0.33
F1-score	0.82	0.39

❑ Decision Tree Model

	Not Canceled = 0	Canceled =1
precision	0.98	0.96
Recall	0.99	0.94
F1-score	0.98	0.95

Comparison Accuracy

	Model	Score
1	Decision Tree Classifier	1.00
2	Logistic Regression	0.97
3	KNN	0.72


Models Comparison
Decision Tree Classifier
KNN
Logistic Regression
Model
0
0.5
1
Score
Score
1
0.95
0.9
0.85
0.8
0.75

Resources

06

1. https://www.kaggle.com/jessemostipak/hotel-booking-demand

2. Exploratory Data Analysis of the Hotel Booking Demand with Python | by Aaqib Qadeer Soomro | Analytics Vidhya | Medium

Thanks ...

Do you have any questions?